Locomotive rotation optimization as basis for efficient rail cargo operation

(B: Professional paper)

Thomas Albrecht[a], Jonatan Gjerdrum[b]
[a] Rail & Transit Solutions, DXC Technology
Bergstraße 2, 01069 Dresden, Germany
[1] E-mail: talbrecht@dxc.com, Phone: +49 (0) 351 47771 45
[b] Planning, Green Cargo
Svetsarvägen 10, 17141 Solna, Sweden

Keywords
Rail cargo, Locomotive Rotation, Mixed-Integer Optimization, Timetable optimization

1 Introduction

To withstand the high pressure of the competition in the rail cargo market, railway undertakings must operate at highest possible efficiency. Optimal resource utilization is an important prerequisite for high efficiency, in particular for locomotives which represent the most expensive resource in rail cargo operations. Most rail cargo operators use computer-aided manual rotation planning, which often does not produce optimal plans because strong variations in traffic demand and irregular traffic in cargo traffic makes planning difficult.

The use of mathematical optimization well integrated into the IT landscape of a rail undertaking can help to reduce the manual planning effort, quantify the number of resources needed to operate a plan and use the available resources in an optimal way.

Green Cargo, the largest Swedish rail cargo undertaking, has been utilizing optimization systems for locomotive planning since the 1990's and introduced a new "Locomotive Optimization System (LOOP)" in 2017. This contribution describes this solution which has been developed by DXC Technology in close cooperation with Green Cargo. It is used for two major problem classes of locomotive planning:

- Tactical (operational) level: This task shall provide a day-to-day plan to produce the actual transport demand. It is performed monthly at Green Cargo.
- Strategic level for yearly planning and strategic scenarios: Here the traffic demand of one template week is assumed to be repeating weekly. Even though this is hardly true on a detailed level for cargo traffic, the assumption can be considered valid for strategic purposes when suitable traffic schemes are chosen, e.g. those of high traffic demand for determining required fleet size. Furthermore, all changes involving adaptations of timetables need to be aligned with the infrastructure manager and are considered as strategic optimization problems.

The article describes the models used for the mathematical optimization of locomotive rotations, how these are applied to practical planning problems and their integration in the planning process and toolchain. Optimization results will be presented in practical case studies based on real-world operating scenarios provided by Green Cargo.

2 Optimization models for tactical planning

2.1 Description of the Optimization problem

Green Cargo uses in the order of 350 locomotives for all its operations in Sweden, Norway and Denmark. These locomotives are optimized to provide the maximum efficiency to the current traffic program. There is no home station to the locomotives which might restrict their potential of usage but some locomotives have special features such as remote control (most), or ability to run on ETCS controlled tracks (few).

The purpose of locomotive planning in general and optimization with respect to scheduling in particular is to satisfy the traffic program requirements with the lowest possible asset number while maintaining robustness and customer satisfaction. Thus, the timetable planning (the most important part of the traffic program in this context) and locomotive scheduling problems are somewhat integrated. Green Cargo applies for timetable slots provided by infrastructure managers (mainly Trafikverket) but maintains a database for many more timetable scenarios which are evaluated in terms of locomotive utilization (as well as other resources such as crew). It is of key importance to apply for the right timetables at the right time with the right pulling power requirements i.e. occasional multiple locomotives. The problem is further complicated by the fact that Green Cargo competes for track capacity with other freight and passenger operators both long haul and commuter traffic and that much of the Swedish network is single-track only. Green Cargo therefore needs to interact closely with infrastructure managers to clarify its requirements both on a long-term and short-term horizon.

The starting point for locomotive optimization is a specific version of the timetable, be it the yearly plan or a monthly update of it. This timetable contains all Green Cargo operated trains and the requirements on locomotives based on so-called task classes which incorporate the special features of the locomotives. A task class comprises one or several locomotive types of similar driving characteristics. Using task classes allows the infrastructure manager to compute train running times with sufficient reliability whereas the railway undertaking still maintains a minimum level of freedom in the use of the actual locomotive types. Beside the task class, the minimum number of locomotives of each task class is also defined in the timetable (i.e. one or two or in rare instances three locomotives).

The required task classes for a train run can change along the journey. Therefore, a train run is split into so-called train legs at operational locations, where a change of task classes is required according to the timetable. To increase planning flexibility, train runs are also split at operational locations, where locomotive changes are allowed and sufficient time is available in the timetable, thereby creating more train legs.

The task of the monthly planning cycle is to identify which locomotive types to use in which number on each train leg (assignment problem) and to find rotations for each locomotive, i.e. the sequence of train legs a locomotive shall run on during the planning period. The monthly planning problem is solved as so-called dated planning problem, i.e. each train run is considered individually for the planning, even if it repeats several times on different days during a month or week. In the process it shall be possible to consider multiple locomotive types at the same time. This optimization problem is decomposed into a three-stage process, see Figure 1:

First, the possible combinations of locomotive types are computed for each train leg (stage 1), then an optimal assignment is searched for in two steps (stage 2.1 and 2.2). The rotation plan is handled as separate decision problem in stage 3.

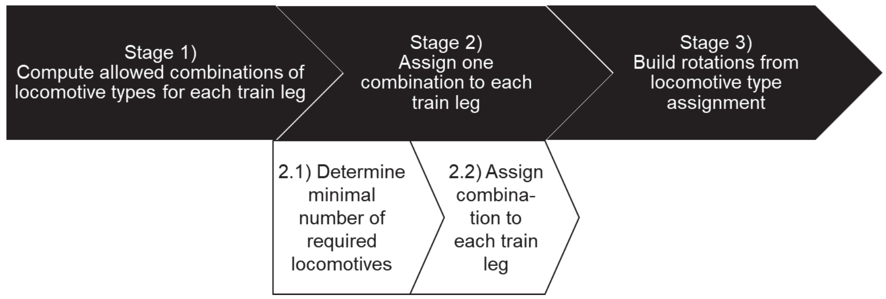

Figure 1: Optimization process architecture

2.2 Finding allowed combinations of locomotive types

For each train leg, the minimal number of required locomotives and their task classes is given. The maximum number of active locomotives depends on the infrastructure used and can be taken from the infrastructure model (typically two or three active locomotives).

For a specific train leg, the use of a specific locomotive type is possible under the following conditions:

1. All power supply systems installed along the train leg must be available on all locomotive types of the active locomotives.
2. One train protection system installed on the track must be available on the locomotive type of the leading locomotive.
3. The use of a locomotive type must be permitted along the entire train leg (The use of some locomotive types might be restricted to certain regions only)

From these constraints, the permitted locomotive types Θ_l for a train leg l are derived. From these permitted locomotive types, the possible combinations $c \in C_l$ for this train leg can be computed. Combinations are only allowed when all locomotive types in a combination are compatible to run in multiple traction with each other. Each combination c consists of a number $\lambda_{c,\theta}$ of locomotives of type θ.

2.3 Modelling approaches for the assignment problem

In a valid solution of locomotive assignment each train leg in the timetable must be served by exactly one of the allowed combinations of locomotive types. Therefore, decision variables $x_{l,c}$ are introduced which take the value 1, if train leg $l \in L$ is served by the combination c of locomotive types and 0 otherwise. The number of locomotives $\lambda_{\theta,l}$ of type θ on leg l can be expressed as

$$\lambda_{\theta,l} = x_{l,c} \cdot \lambda_{c,\theta}$$

For any valid assignment the following constraint must be fulfilled (one combination on each leg):

$$\forall l\colon \sum_{c \in C_l} x_{l,c} = 1$$

Furthermore, it must be guaranteed, that the same locomotive is being used only on train legs, which do not overlap in time. In the literature (see e.g. (Aronsson, M. et al., 2006) and (Giacco, G.L. et al., 2011) for an overview) this constraint is modelled as a multi-commodity network flow problem in a graph using two different approaches (see also Figure 2):

1. *Connection edges*: Each train leg is modelled as a vertex v in the graph. The potential connections of locomotives between train legs are modelled by directed edges e in the graph, i.e. edges are created between each arriving and departing train leg at the same operational location, where the departure time is later than the arrival time (See e.g. Aronsson, M. et al., 2006). Supplementary decision variables $s_{e,\theta}$ are introduced for the number of locomotives of type θ on edge e.
2. *Waiting edges*: Each departure and arrival of a train leg is modelled as a vertex v in a graph. Train legs are modelled as directed edges between these vertices. Furthermore, waiting edges are introduced between consecutive (departure or arrival) vertices at the same operational location o (See e.g. BMWi project, 2005). Supplementary decision variables $s_{e,\theta}$ are introduced for the number of locomotives of type θ on each waiting edge e. When two different trains arrive or depart at the same time, a waiting edge of length 0 is introduced between the vertices, therefore there exists always exactly one incoming waiting edge and one outgoing waiting edge and either a departing train leg edge or an arriving train leg edge for each vertex.

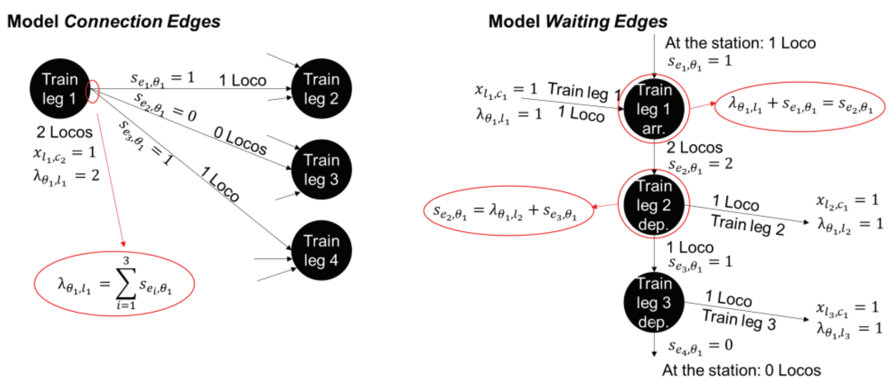

Figure 2: Comparison of the two different graph modelling approaches (assumption: one locomotive type θ_1 only, the combination c_1 has one locomotive, c_2 two locomotives)

For a valid solution, the flow constraint must be fulfilled on each vertex in the graph, where for each locomotive type the number of locomotives on the incoming edges e_{in} must be equal to the number on the outgoing edges e_{out}.

In the *connection edges* model both numbers are equal to the number of locomotives used on train leg l and can be expressed by:

$$\forall \theta \in \Theta, v \in V: \sum_{e_{\text{in},v}} s_{e,\theta} = \lambda_{\theta,l} = \sum_{e_{\text{out},v}} s_{e,\theta}$$

In the *waiting edges* model, for each departure or arrival event of any train leg at an operational location o (modelled as vertex v) the following flow constraint must be fulfilled:

$$\forall \theta \in \Theta, o \in O: \forall v \in V_o: \lambda_{\theta,l_{\text{arr}}} + s_{e_{\text{in},v},\theta} = \lambda_{\theta,l_{\text{dep}}} + s_{e_{\text{out},v},\theta}$$

where either $\lambda_{\theta,l_{\text{arr}}} > 0$ or $\lambda_{\theta,l_{\text{dep}}} > 0$.

The number of decision variables per operational location o and locomotive type is

$l_{\mathrm{dep},o} \cdot l_{\mathrm{arr},o}$ for the *connection edges model* and $2 \cdot (l_{\mathrm{dep},o} + l_{\mathrm{arr},o})$ in the *waiting edges model*.

Additional decision variables can be introduced in the *waiting edges* model to represent the number of locomotives of each type at the start and the end of the considered planning period at each operational location. In the *connection edges* model virtual train runs could be introduced as additional vertices to represent the possible start and end conditions. In practice, this constraint is not used.

During experiments it was shown that both models provide advantages and shortcomings: in the model using *connection edges* constraints for locomotive transfers in stations can easily be considered, e.g. whether there is enough time to couple locomotives at a station. The *waiting edge* model on the contrary requires significantly less decision variables and therefore typically computes in shorter time. Furthermore it allows for the explicit consideration of parking capacity at the stations, although this is currently not considered in the implemented model.

In the previous locomotive planning system used at Green Cargo the dated planning was performed manually based on weekly optimization. Dated optimization as provided by LOOP reduced the time required and the manual work to produce a plan as well as the restrictions for optimization based on manual input. Optimization across various locomotive types also improves the solution obtained.

2.4 Objective function and optimization approach

The objectives for locomotive optimization are manifold and partially contradictory as is often the case when optimizing both on cost and quality. The main objective is to produce the lowest number of locomotives that could satisfy all train legs as well as all constraints. However, extra costs are introduced if there are passive moves of locomotives (a.k.a. deadheading), if there are more locomotive changes in a train on-route or if an expensive locomotive is run when a cheaper one could have been used.

Thus, some of the objectives originally specified by the Green Cargo were:

1. Reduce the number of locomotives needed
2. Reduce the overall distance travelled by all locomotives (compute a plan with the minimal effort for re-positioning locomotives)
3. Consider the running cost of different locomotive types
4. Create robust locomotive rotations (avoid short connections between consecutive locomotive runs, in particular when trains serving different business areas are combined)
5. Avoid overlapping of pre- and post-processing times if possible
6. Try to ensure certain connections between train legs
7. Avoid changing compositions of locomotives (multi traction)

During an intensive experimental phase different approaches of multi-objective optimization (Branke et al., 2008) for producing locomotive rotations have been implemented and the obtained results were examined by the Green Cargo planners. As a compromise between solution quality, computation time and controllability of the solution a combination of lexicographical ordering (which leads to a decomposition of stage 2) and weighted sum has been chosen. In order to satisfy the most important goal of optimization (minimal number of locomotives in a plan) this number is used as objective in a first optimization run without considering any other objective (2.1, see also Figure 1). In all consecutive optimization stages, this minimal number of required locomotives is considered as additional constraint.

The second most important objective is the reduction of operating costs including e.g.

running costs per locomotive types which is considered in a second stage (2.2) as a weighted sum of distance dependent cost per locomotive type on each train leg, considering also different cost for active or passive locomotive usage. For stages 2.1 and 2.2 the waiting edges model is used.

In stage 3, rotations are built. The assignment of locomotive types obtained from stage 2.2 is used as additional constraint in this stage. The fulfilment of the flow constraints (in stages 2.1 and 2.2) guarantees that a valid rotation plan can be built. Decisions on this stage must only be made if more than one locomotive is assigned to a train leg or more than one locomotive is waiting at an operational location which results in a significantly smaller number of decision variables. The objective function is a weighted sum of penalties for unwanted connections within a rotation, e.g. too short connection times, breaking desired connections between consecutive legs of the same train run or between pre-specified pairs of train legs, combining legs of different business areas, uncoupling/ coupling of multi-traction when it can be avoided.

For the locomotive assignment problem, the optimum can only be reached by considering all locomotive types at the same time. Because of the practical assignment of locomotive types to task classes however it is possible to decompose the model into so-called subproblems and thereby reduce computation time significantly (e.g. by treating diesel-hauled and electric locomotive types separately). These subproblems are also created where dedicated fleets shall be used to operate special kinds of traffic (e.g. for postal trains which run at higher speeds than other cargo traffic).

The process of building rotations (stage 3) is executed separately per locomotive type.

3 Strategic optimization problems

Even though a timetable considers a yearly time frame it is practical to extract a specific week and make that week representative of a time period. This is called cyclic planning and for this type of strategic optimization problems, a repeating week is assumed. Cyclic planning is further used to analyze (seasonal) traffic patterns and derive potential measures to control dated planning (by so-called locks, see section 4.2).

The assumption of a cyclic planning problem means that the assignment of locomotives at the end of the considered template week (assignment to train legs and stock in stations) must be equal to the state at the beginning of the template week. In the *waiting edges* model these constraints can be introduced by adding a waiting edge for each operational location starting at the last vertex in the planning period and connecting it to the first vertex of the planning period. In the *connection edges* model connection edges are introduced from an arriving edge to all departing edges regardless of their departure time. If the departure time is earlier than the arrival time of the incoming train the connection edge represents a number of locomotives waiting over the end of the template period.

To compute the number of locomotives needed in cyclic planning, the number of locomotives used on all edges at any time point within the planning interval must be summed up. In the *waiting edges* model this is straightforward, for the connection edges model several special cases have to be considered in particular for very long train runs (Aronsson, M. et al., 2006).

In strategic planning, there are opportunities to create timetables with a better fit to locomotive rotations. Trains could be shifted in time so as to create shorter, more efficient standstills in terminals which in turn can reduce the number of locomotives. The approach presented in (Aronsson, M. et al., 2006) based on a *connection edge model* has been extended for multiple locomotive types and integrated in the LOOP solution. The *waiting* *edge* model is not suitable for this kind of problem as the network topology changes for different timeshift.

4 Practical application

4.1 Implementation in IT solution

Using the described optimization algorithms in practice requires an IT solution which is fully integrated in the IT landscape and thus allows for a high degree of automation of the planning process. LOOP is based on DXC's Rail Cargo Management Solution RCMS (DXC, 2018), which had been equipped with interfaces to the systems used for timetable planning (of the infrastructure manager), crew planning and railway operations management. Infrastructure data is imported from the infrastructure manager system. Different validity periods of infrastructure data are modelled in LOOP in order to consider (future) changes in network topology for simulation of scenarios. LOOP is the leading system for all data on locomotive types which is used to assure the compatibility between locomotive and tracks and locomotives of different types among each other. These comprehensive network and locomotive models allow a high degree of automation of the planning process.

The rotation creation and optimization process use the standard RCMS scenario technology: The timetable is imported into a so-called timetable scenario, for which different resource scenarios can be created which contain the (interative improvements of the) locomotive rotations. The planning solutions obtained by the optimizer are presented to the planners in different GANTT charts. Here the planners can analyze the results and change the plans interactively. By introducing so-called locks between one or multiple consecutive train legs they can create input for a next optimizer run. There are so-called hard-locks on connections, which must not be broken by the optimizer, and soft-locks, which can be broken at the cost of a penalty only. The planning results are also presented in tabular format which can be exported for further analysis.

If any of the optimization runs does not find a suitable solution (in stage 1 of the optimization process), LOOP provides different views to analyze the root causes of the infeasibility and includes specific optimization problems to identify infeasible legs.

The solution is built in Java and incorporates CPLEX as solver for the different optimization models. The LOOP system is in full productive use at Green Cargo since summer 2017.

4.2 Planning process

In the yearly process, the aim is to provide a template for the coming year and make sure that the locomotive fleet is sufficient to enable the traffic program or to propose changes to the fleet sizes. At the same time, a number of productivity targets are set and changes to timetables are proposed. The locomotive planner is both a stakeholder and a support person in this process. Timing between arriving trains and departing trains in a station is crucial to create locomotive rotation plans that are efficient and robust. Therefore, several train planning related strategic optimization methods are tested to reduce the number of locomotives required whereas maintaining low effects on the traffic program: timetable shift, passive moves and remove trains in a pre-specified set. These strategic optimization methods based on a tight integration between timetable, traffic program and locomotive optimization are foreseen to have positive implications on the future locomotive plans. They are currently under test at Green Cargo.

The implemented yearly timetable is the basis for the monthly plan. The monthly process is repetitive in nature and mainly reacts to factors such as (mainly smaller) variations in business volumes, track work and sometimes locomotive maintenance. Typically, the planner imports all relevant trains from the timetable system and iteratively optimizes the plan for the month of interest until a sufficiently good match between trains and locomotive rotations is found. Manual input based on expert knowledge is made both in the timetable system and in LOOP, typically by balancing the number of trains to and from stations, e.g. by empty runs, or by controlling the optimizer through parameter settings or restricting which locomotive turns are permitted.

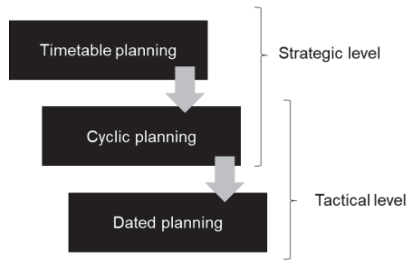

Figure 3: Overview over planning levels

5 Case studies

In order to illustrate the effects of the different models on computation time they have been run in different real-world scenarios. The results of these experiments (one run per scenario) are given in Table 1. The small scenario comprises one part of Green Cargo traffic which is run with a dedicated locomotive fleet. The big scenario contains all electrically-hauled traffic which is not run by dedicated fleets. The connection edges model is only used in cyclic planning. In dated planning the plan for a full month (31 days) is computed with the waiting edges model.

The number of train legs and the obtained number of locomotives gives an indication of the problem size. The small difference between the connection edges model and the waiting edges model in the number of train legs is due to the fact, that some hard-lock constraints are not considered in the connection edge model. The number "non-zeros" refers to the non-zero elements in the problem matrix after pre-processing by CPLEX. It is a good indicator on problem complexity and computational effort (Hill, F. et al., 1984). It can be seen that this number is approximately 20 times higher in the big scenario with the connection edges model compared to the waiting edges model. The computation time is compared for the first step in the locomotive assignment process, i.e. the computation of the minimal number of required locomotives and the proof of optimality by CPLEX. All experiments ran on a 4-core Intel Server with 2.6 GHz processor and 8 GB RAM. It can be seen that the small problems solve immediately regardless of the used model, but there are significant differences in the computation time of the big problem. The connection edges model takes very long to compute a valid solution in the cyclic problem. The waiting edges model performs significantly better in the cyclic problem and even the computation time for the big dated planning problem with three times more train legs is shorter than the one required by the connection edges model for cyclic planning. The table also shows the total computation times of the rotation building process (including all stages 2.1, 2.2 and 3).

The computation times are acceptable from the viewpoint of Green Cargo for this kind of planning problems. Even in the dated planning they allow for several planning iterations per working day. It has also been shown that the waiting edge model is able to compute a reliable lower bound for the number of needed locomotives within one CPLEX iteration, i.e. within a few seconds. This property is used in practice to speed up dated optimization of the big problem to less than one hour.

Table 1: Computational results (Computation times are given in h:mm:ss)

	Cyclic planning				Dated planning	
Scenario	Small	Small	Big	Big	Small	Big
Model used	Conn. edges	Waiting edges	Conn. edges	Waiting edges	Waiting edges	Waiting edges
Train legs	70	70	1 746	1 671	290	6 926
Non-zeros	1 292	466	1 119 454	57 226	1 887	238 159
Locomo-tives	9	9	161	164	9	169
Comp. time assign-ment	<0:00:01	<0:00:01	5:00:14	0:02:17	<0:00:01	0:30:14
Comp. time full	-	<0:00:01	-	0:03:32	0:00:01	1:21:13

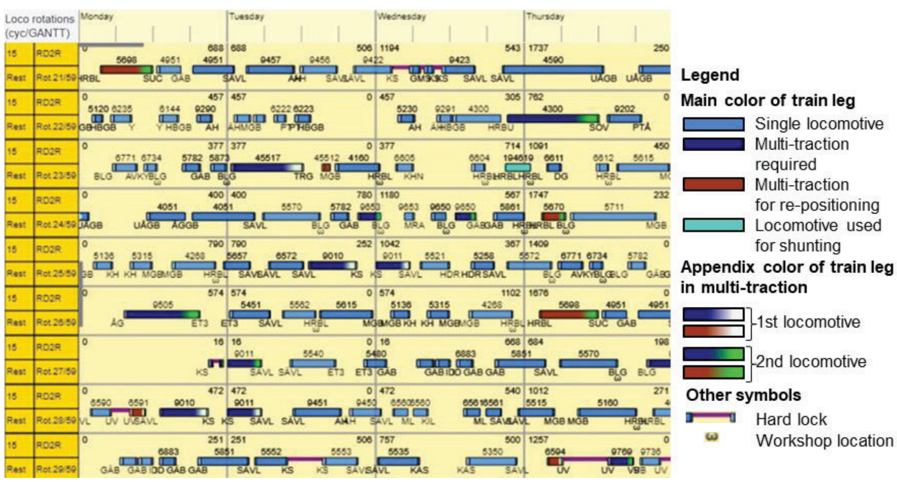

Figure 4: Excerpt of a GANTT view of the locomotive rotations in LOOP for a cyclic plan

Figure 4 shows an excerpt of the main GANTT chart displaying results of the cyclic optimization of the big scenario. One locomotive rotation is displayed per row. In the left column (orange, the rotation header), locomotive type (here: RD2R), and rotation number within a cycle are given (e.g. Rot. 25/59 is week 25 in a rotating cycle of 59 weeks in length). It should be noted that the planning cycle length is not part of the planning objectives. In the right (yellow shaded part) of the figure, the sequence of train legs planned for the locomotive is displayed. The different colors, shadings and symbols allow for quick check of the plan efficiency and correctness by the planners.

6 Conclusions and further Research

The introduction of the Locomotive Optimization System LOOP allowed Green Cargo to reduce the number of locomotives needed to operate the timetable by increasing the so-called locomotive productivity, i.e. the distance travelled in commercial operation per locomotive. Moreover, process and integration development increased planning speed and improved the solution quality. The perceived key benefits from a Green Cargo planning process perspective are a lower amount of user restrictions allows for better optimization, dated optimization and planning of multiple locomotive types at the same time allow for less manual work and a better integration with the timetable and network data that maintains feasibility of locomotive types on the track network.

A further improvement of locomotive usage is foreseen with a tighter integration of LOOP with the surrounding planning processes and tools:

1. A tighter integration with load and timetable planning would allow to fix the locomotive categories and required minimal number later in the planning process allowing for improved usage of the most recent powerful and multi-purpose locomotives.
2. An integration with crew scheduling is challenging in particular in cargo operation and under the Swedish geography with long travelled distances. But it could be highly beneficial as it would allow reducing the number of manual constraints (locks) that are introduced today by the planners to consider driver constraints during locomotive planning and might thereby reveal new options for driver and locomotive rotations.
3. A tighter coupling with maintenance planning and workshop task scheduling is particularly interesting when workshop capacities are limited and/ or new methods and processes for preventive locomotive maintenance are introduced.
4. The application of LOOP for operational planning and adaptation of locomotive rotations considering real-time information on timetable deviations would make it possible to profit from the optimization capabilities during unplanned events.

References

Aronsson, M., Kreuger, P. and Gjerdrum, J.: *An efficient MIP Model for Locomotive Scheduling with Time Windows*, ATMOS 2006

Giacco, G.L., D'Ariano, A., Pacciarelli, D.: Rolling stock rostering optimization under maintenance constraints, In: *2nd International Conference on Models and Technologies for Intelligent Transportation Systems*, Leuven, Belgium, 2011.

BMWi-Projekt Innovationen für Gleisanschlussverkehre, 2005

Branke, J., Deb, K., Miettinen, K., Slowinski, R. (eds.): *Multiobjective Optimization, Interactive and Evolutionary Approaches*, Springer, LNCS 5252, 2008.

DXC RCMS Product Sheet: http://www.dxc.technology/travel_and_transportation/offerings/43437/81345-rail_cargo_management_system, last access: 01.02.2018

Hill, P.E.., Murray, W., Saunders, M.A. and Wright, M.H.: *Sparse matrix methods in Optimization.* SIAM J. Sci. Stat. Comput., Vol. 5, No. 3, 562-589, 1984